Process control and quality, 1 (1990) 41-51
Elsevier Science Publishers B.V., Amsterdam

A theoretical basis for the use of principal component models for monitoring multivariate processes

B.M. Wise, N.L. Ricker *, D.F. Veltkamp and B.R. Kowalski
Center for process Analytical Chemistry, BG-10, and Department of Chemical Engineering, BF-10, University of Washington, Seattle, WA 98195

(Received March 16, 1990, accepted June 4, 1990)

ABSTRACT Although principal components analysis (PCA) has shown some utility as a multivariate statistical process control (MSPC) tool, a theoretical basis for its use with dynamic process data has been lacking. It is shown here that, for processes with more measurements than states, proper application of PCA can facilitate the process monitoring and fault detection problem. When the PCA model is accurate, variations in the process states appear primarily as variations in PCA scores, while noise mainly affects the residuals. This allows one to consider on the the noise properties when deriving error-detection limits for the PCA residuals. In particular, the process dynamics need not be considered explicitly. We also show how an arbitrary stat-space model can be transformed so that its states are directly related to the PCA scores.

Keywords: principal components – statistical process control

Process control and quality, 1 (1990) 41–51
Elsevier Science Publishers B.V., Amsterdam

A theoretical basis for the use of principal component models for monitoring multivariate processes

B.M. Wise, N.L. Ricker*, D.F. Veltkamp and B.R. Kowalski

Center for Process Analytical Chemistry, BG-10, and Department of Chemical Engineering, BF-10, University of Washington, Seattle, WA 98195 (U.S.A.)

(Received March 16, 1990; accepted June 4, 1990)

Abstract

Although principal components analysis (PCA) has shown some utility as a multivariate statistical process control (MSPC) tool, a theoretical basis for its use with dynamic process data has been lacking. It is shown here that, for processes with more measurements than states, proper application of PCA can facilitate the process monitoring and fault detection problem. When the PCA model is accurate, variations in the process states appear primarily as variations in the PCA scores, while noise mainly affects the residuals. This allows one to consider only the noise properties when deriving error-detection limits for the PCA residuals. In particular, the process dynamics need not be considered explicitly. We also show how an arbitrary state–space model can be transformed so that its states are directly related to the PCA scores.

Keywords: principal components – statistical process control

INTRODUCTION

Because of recent advances in process sensors and data acquisition systems, today's chemical processes are becoming better instrumented. In many cases, this instrumentation provides redundant data, i.e., some of the measurements are highly correlated. Principal components analysis (PCA) can express the essential information contained in these measurements in terms of a relatively small number of "latent variables". The utility of doing this has been demonstrated in previous work [1–4], in which PCA has been applied to chemical process data, and multivariate statistical process control (MSPC) methodology has been developed.

Recently, Ricker [5] has shown that when the process can be considered to be at steady-state, PCA is equivalent to a special type of Kalman filter. Until now, however, a theoretical basis for the use of PCA with *dynamic* process data has been lacking. The main contribution of this paper is to show when it is appropriate to apply MSPC techniques to such data, and how the results should be interpreted.

* To whom correspondence about this paper should be addressed.

0924-3089/90/$03.50

PRINCIPAL COMPONENTS ANALYSIS

We consider an m by n data matrix [1], $\mathbf{X}$, in which each row, $\boldsymbol{x}_i^{\mathrm{T}}$, is a sample of n variables taken at a specific time. PCA decomposes such a matrix into the sum of the products of n pairs of vectors [6,7]. Each pair consists of a vector in $\mathscr{R}^n$ called the loadings, $\boldsymbol{p}_i$, and a vector in $\mathscr{R}^m$ referred to as the scores, $\boldsymbol{t}_i$. Thus $\mathbf{X}$ can be written as

$$\mathbf{X} = \boldsymbol{t}_1\boldsymbol{p}_1^{\mathrm{T}} + \boldsymbol{t}_2\boldsymbol{p}_2^{\mathrm{T}} + \cdots + \boldsymbol{t}_n\boldsymbol{p}_n^{\mathrm{T}} \tag{1}$$

The matrix of loadings vectors, $\mathbf{P}$, forms an orthogonal basis for the space spanned by $\mathbf{X}$ and the individual $\boldsymbol{p}_i$ are the eigenvectors of the scatter matrix of $\mathbf{X}$, defined as:

$$\text{scatter}(\mathbf{X}) = \frac{\mathbf{X}^{\mathrm{T}}\mathbf{X}}{m-1} \tag{2}$$

Thus

$$\text{scatter}(\mathbf{X})\boldsymbol{p}_i = \lambda_i\boldsymbol{p}_i \tag{3}$$

where λ_i is the eigenvalue associated with the eigenvector $\boldsymbol{p}_i$. If $\mathbf{X}$ has been *mean-centered* (i.e., each variable is scaled to have zero mean) the scatter matrix is equal to the covariance matrix. If $\mathbf{X}$ has been mean-centered and scaled to unit variance (autoscaled), the scatter matrix is equivalent to the correlation matrix. The loadings vectors are often referred to as principal components, or as "latent variables" (particularly in Partial Least Squares calibration) because they are linear combinations of the original variables that together explain large fractions of the information in the $\mathbf{X}$ matrix. Each of the $\boldsymbol{t}_i$ is simply the projection of $\mathbf{X}$ onto the basis vector $\boldsymbol{p}_i$:

$$\boldsymbol{t}_i = \mathbf{X}\boldsymbol{p}_i \tag{4}$$

The value of each λ_i is equal to the variance in the data set associated with the vector direction $\boldsymbol{p}_i$. In fact, fraction variance in direction $\boldsymbol{p}_i$ is given by

$$\boldsymbol{p}_i = \lambda_i / \sum_{i=1}^{n} \lambda_i \tag{5}$$

PCA is very closely related to the Singular Value Decomposition (SVD) [6], by which a data matrix, $\mathbf{X}$, is decomposed as

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}} \tag{6}$$

Here, the columns of $\mathbf{V}$ are equal to the PCA eigenvectors ($\boldsymbol{p}_i$) and Σ is a diagonal matrix containing the singular values, which are equal to the square roots of the eigenvalues of scatter($\mathbf{X}$).

Once the eigenvectors have been determined *via* PCA or SVD, selected samples can be projected onto the eigenvectors. These projections are commonly referred to as "scores plots" and are often useful for showing the relationships between the samples. (In this context, "sample" refers to the vector of process measurements taken at a specific instant in time.) Plots can be made of the projections of the sample onto a single eigenvector versus sample number (or time), or onto the plane formed by two eigenvectors. A projection of the samples onto the two eigenvectors associated with the largest eigenvalues depicts the largest amount of information about the relationship between the samples that can be shown in two dimensions using a linear projection. It is for this reason that PCA is often used as a pattern recognition and sample classification technique.

Plots of the coefficients of the eigenvectors themselves, known as "loadings plots", show the relationships between the original variables in the data set. Correlations between variables are apparent in these loadings plots.

When a data set is subjected to PCA, it is often found (and it is generally the objective) that only the first few eigenvectors are associated with systematic variation in the data, and the remaining eigenvectors are associated with "noise". Noise in this case refers to uncontrolled process and instrumental variations arising from random influences. PCA models are formed by retaining only the eigenvectors that describe systematic variations in the data. Determina-

[1] In this paper, all matrices are upper-case, boldface variables. All vectors are lower-case, boldface italic variables and are *column* vectors. A superscript "T" indicates the transpose. Thus, $\boldsymbol{x}^{\mathrm{T}}$ is a *row* vector.

tion of the proper number of eigenvectors can be done by cross-validation [8] or other techniques, such as comparison of the distribution of eigenvalues to the distribution expected for random variables [9].

Once the PCA model is formed, new data can be viewed as projections onto single eigenvectors or onto the plane formed by a pair of eigenvectors. The "goodness of fit" between new data and the model can be monitored by calculation of the "residual". For a reduced-order model, $\mathbf{P}_q$, (where $\mathbf{P}_q$ is made up of the first q eigenvectors in $\mathbf{V}$) and a new data matrix, $\mathbf{X}$, the residual matrix $\mathbf{R}$ is given by

$$\mathbf{R} = \mathbf{X}\left(\mathbf{I} - \mathbf{P}_q \mathbf{P}_q^{\mathrm{T}}\right) \tag{7}$$

For any sample, $\boldsymbol{x}_i^{\mathrm{T}}$, in $\mathbf{X}$, the sum-of-squared-residuals (which can also be thought of as the squared length of the residual vector, $\boldsymbol{r}_i^{\mathrm{T}}$ in n dimensional space) is

$$Q = \boldsymbol{r}_i^{\mathrm{T}} \boldsymbol{r}_i = \boldsymbol{x}_i^{\mathrm{T}}\left(\mathbf{I} - \mathbf{P}_q \mathbf{P}_q^{\mathrm{T}}\right)\boldsymbol{x}_i \tag{8}$$

The value of the scalar, Q, is a measure of the "goodness of fit" of the new sample, $\boldsymbol{x}_i^{\mathrm{T}}$, to the model $\mathbf{P}_q$.

STATE-SPACE MODEL OF A DYNAMIC PROCESS

Consider a linear, time-invariant (LTI), discrete, state–space process model of the form:

$$\boldsymbol{x}(k+1) = \boldsymbol{\Phi}\boldsymbol{x}(k) + \boldsymbol{\Gamma}\boldsymbol{u}(k) + \boldsymbol{v}(k) \tag{9}$$

$$\boldsymbol{y}(k) = \mathbf{C}\boldsymbol{x}(k) + \mathbf{D}\boldsymbol{u}(k) + \boldsymbol{e}(k) \tag{10}$$

where $\boldsymbol{x}(k)$ is the state vector (in $\mathscr{R}^n$) at sampling period k, $\boldsymbol{u}(k)$ is the input vector (in $\mathscr{R}^r$), and $\boldsymbol{y}(k)$ is the output measurement vector (in $\mathscr{R}^p$). The vector $\boldsymbol{v}(k)$ represents the state noise or disturbance inputs; $\boldsymbol{e}(k)$ is measurement noise, which, for periods of "normal" operation is assumed to be random with zero mean. The $\boldsymbol{\Phi}$, $\boldsymbol{\Gamma}$, $\mathbf{C}$, and $\mathbf{D}$ matrices are assumed to be constant. Note that eqn. (9) is a recursive relationship that allows one to calculate the state vector at sampling period $k+1$ in terms of the states and inputs at period k. Equation (10) shows how the measurements, $\boldsymbol{y}(k)$, are related to the states, inputs, and measurement noise. Many references concerning the state–space formalism are available from literature [10–12].

For the purposes of this paper, we assume that $\mathbf{D} = \mathbf{0}$, which implies that changes in the inputs, $\boldsymbol{u}(k)$, do not affect the measurements, $\boldsymbol{y}(k)$, instantaneously. Due to the delay caused by sampling, this is a realistic assumption of most chemical processes. If, however, $\mathbf{D}$ is known (and non-zero), the effect of the term $\mathbf{D}\boldsymbol{u}(k)$ in eqn. (10) can be subtracted from $\boldsymbol{y}(k)$, and the methods described below can still be applied in a straightforward manner.

Note that although the *number* of state variables, n, is a fundamental property of a process [2], the *coordinate system* defining the numerical values of the states may be chosen arbitrarily. We exploit this property of the model in the next section.

Immediately observable states

We define the "immediately observable" states as the subset of the n state variables that can be estimated from a *single* sample of the p outputs. From an examination of eqn. (10), it is easy to see that the number of immediately observable states is equal to the rank of $\mathbf{C}$. Let rank $(\mathbf{C}) = q$ and note that $q \leq \min(n, p)$.

One can transform the state–space model given in eqns. (9) and (10) to a form in which the immediately observable states are easy to calculate. We first perform a singular-value decomposition on the $\mathbf{C}$ matrix:

$$\mathbf{C} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}} \tag{11}$$

In this case, $\mathbf{U}$ is p by p, $\mathbf{V}$ is n by n, and Σ has the form:

$$\Sigma = \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{for } q < p \text{ and } q < n \tag{12a}$$

$$\Sigma = \begin{bmatrix} \mathbf{S} \\ \mathbf{0} \end{bmatrix} \quad \text{for } q < p \text{ and } q = n \tag{12b}$$

$$\Sigma = \begin{bmatrix} \mathbf{S} & \mathbf{0} \end{bmatrix} \quad \text{for } q = p \text{ and } q < n \tag{12c}$$

[2] This is usually called the system "order".

where $\mathbf{S}$ is a q by q diagonal matrix of singular values. Define a new matrix, $\mathbf{Q}$, as follows:

$$\mathbf{Q} = \mathbf{V}\begin{bmatrix} \mathbf{S}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \quad \text{for } q < n \tag{12d}$$

$$\mathbf{Q} = \mathbf{V}\mathbf{S}^{-1} \quad \text{for } q = n \tag{12e}$$

where in eqn. (12d), the matrix in brackets is n by n, and $\mathbf{I}$ is an identity matrix of size $n - q$. Note that in eqn 12(e) $\mathbf{Q}$ is defined such that $\mathbf{Q}^{-1}$ exists. We can now define a new coordinate system for the state variables, which is related to the original one according to: $\boldsymbol{x}_r(k) = \mathbf{Q}^{-1}\boldsymbol{x}(k)$, where $\boldsymbol{x}_r$ is the state vector in the new (rotated) coordinate system. Substituting $\boldsymbol{x}(k) = \mathbf{Q}\boldsymbol{x}_r(k)$ into eqns. (9) and (10), we obtain (for $\mathbf{D} = \mathbf{0}$):

$$\boldsymbol{x}_r(k+1) = \boldsymbol{\Phi}_r\boldsymbol{x}_r(k) + \Gamma_r\boldsymbol{u}(k) + \mathbf{Q}^{-1}\boldsymbol{v}(k) \tag{13}$$

$$\boldsymbol{y}(k) = \mathbf{C}_r\boldsymbol{x}_r(k) + \boldsymbol{e}(k) \tag{14}$$

where

$$\boldsymbol{\Phi}_r = \mathbf{Q}^{-1}\boldsymbol{\Phi}\mathbf{Q} \tag{15}$$

$$\Gamma_r = \mathbf{Q}^{-1}\Gamma \tag{16}$$

$$\mathbf{C}_r = \mathbf{U}\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{for } q < p \text{ and } q < n \tag{17a}$$

$$\mathbf{C}_r = \mathbf{U}\begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \quad \text{for } q < p \text{ and } q = n \tag{17b}$$

$$\mathbf{C}_r = \mathbf{U}[\,\mathbf{I} \quad \mathbf{0}\,] \quad \text{for } q = p \text{ and } q < n \tag{17c}$$

In eqn. (17), the matrix in brackets is p by n in all cases, and the identity matrix is of dimension q. Note that the new state–space model has the same form as that in eqns. (9) and (10), and the input and output variables are also identical. Thus, the only effect of the change of coordinates is in the numerical values of the states.

As is apparent from eqn. (14), only the first q states in $\boldsymbol{x}_r$ have a direct effect on the outputs. Let $\boldsymbol{x}_q(k)$ be the vector of the first q variables in $\boldsymbol{x}_r(k)$, i.e., the immediately observable states. Then eqn. (14) can be re-written as:

$$\boldsymbol{y}(k) = \mathbf{P}_q\boldsymbol{x}_q(k) + \boldsymbol{e}(k) \tag{18}$$

where $\mathbf{P}_q$ is an orthonormal matrix defined as:

$$\mathbf{P}_q = \mathbf{U}\begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \text{for } q < p \tag{19a}$$

$$\mathbf{P}_q = \mathbf{U} \quad \text{for } q = p \tag{19b}$$

Given a sample of the outputs, $\boldsymbol{y}(k)$, we can use eqn. (18) to estimate $\boldsymbol{x}_q(k)$ as follows:

$$\hat{\boldsymbol{x}}_q(k) = \mathbf{P}_q^{\mathrm{T}}\boldsymbol{y}(k) \tag{20}$$

The corresponding estimate of the measurement noise is:

$$\begin{aligned} \hat{\boldsymbol{e}}(k) &= \boldsymbol{y}(k) - \hat{\boldsymbol{y}}(k) \\ &= \left(\mathbf{I} - \mathbf{P}_q\mathbf{P}_q^{\mathrm{T}}\right)\boldsymbol{y}(k) \end{aligned} \tag{21}$$

where

$$\begin{aligned} \hat{\boldsymbol{y}}(k) &= \mathbf{P}_q\hat{\boldsymbol{x}}_q(k) \\ &= \mathbf{P}_q\mathbf{P}_q^{\mathrm{T}}\boldsymbol{y}(k) \end{aligned} \tag{22}$$

As discussed, e.g. by Ricker [13], eqn. (20) provides an optimal estimate of the states in the sense that it minimizes $\hat{\boldsymbol{e}}^{\mathrm{T}}(k)\hat{\boldsymbol{e}}(k)$, i.e., it is the solution to the linear-least-squares problem. Note that when $q = p$, $\mathbf{P}_q\mathbf{P}_q^{\mathrm{T}} = \mathbf{I}$, and the estimated measurement noise will be zero, regardless of the values of the outputs. In this case, the estimated measurement noise cannot provide useful diagnostic information about the system. Otherwise, however, we can compare the current value of $\hat{\boldsymbol{e}}(k)$ to our expectations (based on a statistical analysis of past process behavior). As we shall show in the next section, this forms the basis for PCA monitoring of the process. The most favorable situation occurs when $p \gg q$, and, in particular, when $p \gg q = n$. In this case, a fault in a single measurement is most likely to show up in the residuals rather than in the estimated states.

The state-estimation approach described above is quite different from that used in the development of the Kalman filter (see, e.g. [5,10]). The goal of the Kalman filter is to obtain estimates of all n of the states. In the present case, this can only happen when $q = n$. Furthermore, the Kalman filter is designed to minimize the squared error in the state estimates (*not* the estimated measurement noise). To do so, the filter compensates for the statistical characteristics of both the measurement noise, $\boldsymbol{e}(k)$, *and* the state disturbances, $\boldsymbol{v}(k)$. Its main disadvantage is that it requires a complete dynamic model of the system (i.e., the $\boldsymbol{\Phi}$ and Γ matrices in addition to the $\mathbf{C}$ matrix), and

one must specify the expectation of $\boldsymbol{e}(k)$ and $\boldsymbol{v}(k)$. The large amount of required information is difficult to obtain for a chemical process of typical complexity. Equations (20)–(22), on the other hand, are essentially a simpler "filter", which can be designed based on a knowledge of **C** only.

THE BASIS FOR PCA MONITORING

Imagine that we have an immediately observable process that can be modeled by eqns. (9) and (10). Let n be the number of states needed to describe the process dynamics. Assume that we have collected m "calibration" samples, $\boldsymbol{y}^{\mathrm{T}}(k)$, of the p outputs, forming the m by p matrix, **Y** (with $m \geq p$). Then according to eqn. (10), the relationship of the sampled outputs to the states and measurement noise is:

$$\mathbf{Y} = \mathbf{X}\mathbf{C}^{\mathrm{T}} + \mathbf{E} \tag{23}$$

where **X** is an m by n matrix of state variables (in which each row is the vector of state variables at a particular sampling period), and **E** is an m by p matrix of measurement noise signals. Both **X** and **E** are unknown, and in the general case, $\mathrm{rank}(\mathbf{X}) = n$ and $\mathrm{rank}(\mathbf{E}) = p$.

If we perform a PCA decomposition on **Y**, retaining q latent variables, we obtain:

$$\mathbf{Y} = \mathbf{T}_q\mathbf{P}_q^{\mathrm{T}} + \mathbf{E}_{p-q} \tag{24}$$

Comparing this result to eqns. (14) and (23), we see that PCA can be interpreted as a state–space model, with $\boldsymbol{T}_q$ representing estimates of the q immediately observable state variables (in a particular coordinate system) at each sampling period, and $\mathbf{P}_q$ taking the place of the **C** matrix. Since $\mathbf{P}_q$ is orthonormal, PCA automatically gives us a state–space model in the form of eqn. (14). Note, however, that $\mathrm{rank}(\mathbf{E}_{p-q})$ equals $p-q$. In other words, unless the "true" measurement noise, **E**, is also rank $p-q$, the PCA estimate will be biased. This may or may not be a problem in practice, depending on the nature of the experiments used to obtain **Y**, as discussed in the next section.

Let us assume for the moment that $\mathbf{P}_q$ is an accurate representation of the true **C** matrix (which implies that we know the number of immediately observable states, q). The PCA residual for a given output sample, $\boldsymbol{y}^{\mathrm{T}}(k)$, is defined as:

$$\boldsymbol{r}(k) = \left(\mathbf{I} - \mathbf{P}_q\mathbf{P}_q^{\mathrm{T}}\right)\boldsymbol{y}(k) \tag{25}$$

Thus the PCA residuals are identical to the estimated measurement noise as defined in eqn. (21). In other words, if the model is correct, the residuals are a function of $\boldsymbol{e}(k)$ only, regardless of the system dynamics. Since we have stipulated that, under normal conditions, $\boldsymbol{e}(k)$ is random with zero mean (and uncorrelated with previous values of $\boldsymbol{e}$), we can use well-established statistical methods to monitor $\boldsymbol{r}(k)$ and verify that it indeed satisfies this condition.

OBTAINING AN ACCURATE ESTIMATE OF $\mathbf{P}_q$

The success or failure of PCA monitoring depends, to a large extent, on the accuracy of the model, i.e., the matrix $\mathbf{P}_q$. Normally [3], this matrix must be estimated from the calibration data, **Y**, as outlined in the previous section. It is clear that the most favorable situation [4] occurs when the term $\mathbf{X}\mathbf{C}^{\mathrm{T}}$ in eqn. (23) is large relative to the noise, **E**. If the system is controllable (see, e.g. [10]), then it is possible to carry out an experiment in which the input variables, $\boldsymbol{u}(k)$, are varied so as to achieve a good signal-to-noise ratio for all n state variables.

On the other hand, if $\boldsymbol{u}$ is allowed to vary "naturally", or if the system includes states that are accessible only through the disturbance vector, $\boldsymbol{v}$, then it will not be possible to guarantee adequate excitation of all n states. For example,

[3] The exception is when an accurate theoretical model of the system is available, but this is rare.

[4] In practice, however, good estimates of $\mathbf{P}_q$ may be obtained even for very noisy systems provided that the noise is uncorrelated, i.e., when the singular values of $\mathbf{X}\mathbf{C}^{\mathrm{T}}$ are large relative to the singular values of **E**. This is the case in the example application described at the end of the paper.

if calibration data were collected during a period when v was relatively small, the resulting estimate of $\mathbf{P}_q$ might not reflect the influence of all the states. If this $\mathbf{P}_q$ were then used to filter new data, as in eqn. (21), and v suddenly became large, the value of $\hat{e}(k)$ might signal a fault. This may or may not be desirable depending upon the situation. If the purpose of the monitoring were to detect disturbances, such a model would be effective. If the objective were to detect sensor failure and fundamental process changes, however, the model could give false alarms when disturbances occurred.

USING THE PCA RESIDUALS

Once an estimate of $\mathbf{P}_q$ is obtained, it can be used to calculate residuals as shown in eqn. (25), and the sum-of-squared-residuals, Q, as shown in eqn. (8). Jackson and Mudholkar [14] used the results of Jenson and Solomon [15] to show that approximate confidence limits can be calculated for Q, provided that all the eigenvalues of covariance ($\mathbf{Y}$) are known. In practice, the underlying distribution of the residuals of individual variables can vary substantially from Gaussian without affecting the results.

The confidence limit on Q can be calculated as:

$$Q_\alpha = \frac{\Theta_1}{h_0}\left[\frac{c_\alpha\sqrt{2\Theta_2 h_0^2}}{\Theta_1} + 1 + \frac{\Theta_2 h_0(h_0-1)}{\Theta_1^2}\right] \tag{26}$$

where

$$\Theta_i = \sum_{j=q+1}^{n} \lambda_j^i \qquad \text{for } i = 1, 2, 3 \tag{27}$$

and

$$h_0 = 1 - \frac{2\Theta_1\Theta_3}{3\Theta_2^2} \tag{28}$$

In eqn. (26), c_α is the normal deviate corresponding to the upper $(1-\alpha)$ percentile. Equation (27) simply states that the Θ_i are equal to the sum of the eigenvalues for the eigenvectors not used in the model, taken to the i^{th} power. Note, however, that this result was derived assuming random errors of mean zero, etc. Autocorrelated data would certainly affect this result. In practice, the present authors have found that the application of eqns. (26)–(28) can lead to erroneous results when there is one principal component not retained in the PCA model with an eigenvalue that is much larger than the remaining eigenvalues. When this occurs, the value of h_0 can be less than zero and the Q_α limits are clearly wrong. In this case, however, the principal component (PC) responsible for the large eigenvalue should probably be retained in the model.

The variables responsible for large Q values can often be found through normal statistical process control methods which track single variables. However, there are instances when these methods fail to detect systematic changes in the process or its sensors because the values of the individual variables have not gone "out of bounds" but have instead just become uncorrelated (or changed their correlation) with the remaining variables.

There are several methods for determining the source of the large Q values in this case. The simplest method is to calculate the column norm (the sum of squares over the variables, instead of over the samples, as is done to calculate Q values) for the residuals matrix for the samples with large Q values. Generally, the perturbed variables will have the largest residuals. This method can fail, however, because it does not account for the fact that the average size of the residual is different for different variables in a data set. What is needed is confidence limits on the residuals, so that it is possible to determine whether the residuals are abnormally large or noisy. The calculation of these confidence limits is shown below.

In other cases, the factors responsible for large values of Q can be found by subjecting the matrix of $\boldsymbol{r}_i$ vectors to PCA (off-line). This determines the major source of variation in the data not accounted for by the original PCA model. Typically the variable(s) with the largest (absolute value) coefficient in the first eigenvector from the residuals matrix will be the varia-

ble(s) responsible for the deviation from the PCA model.

While the Q statistic offers a way to test whether the process data has shifted outside the normal operating space, there is also a statistic that provides an indication of unusual variability within the space. This is Hotelling's T^2 statistic [16]. The value of T^2 for a sample is equal to the sum of squares of the scores (adjusted to unit variance) on each of the principal components in the model. The statistical confidence limits for the values for T^2 can be calculated by means of the F-distribution as follows

$$T^2_{q,m,\alpha} = \frac{q(m-1)}{m-q} F_{q,m-q,\alpha} \tag{29}$$

Here, m is the number of samples in the data set used to calibrate the PCA model, q is the number of principal component vectors retained, and α corresponds to the standard normal deviate. In practice, the authors have found that the T^2 statistic is not as useful as the Q statistic when dealing with process data. Typically, individual process variables will be "out of bounds" before an error is indicated by T^2. We include it here, however, for the sake of completeness, and because there may be processes outside of the authors' experience for which it is useful.

The residual variance for each variable can be calculated for the PCA model. It can be shown that for a given data set, $\mathbf{X}$, with a full set of PCA vectors, $\mathbf{P}$ (of which q are retained), and eigenvalues λ, then the variance of the residual for the j^{th} variable is

$$s_j^2 = \sum_{i=q+1}^{n} p_{ij}^2 \lambda_i \tag{30}$$

where p_{ij} is the loading of the j^{th} variable in the i^{th} principal component (PC). For the data matrix from which the model was obtained this relationship will be exact. If, however, it is assumed that the eigenvalues of all the PCs not retained in the model are equal (a common assumption), then the variance in the residual of the j^{th} variable can be estimated from

$$\hat{s}_j^2 = \left(\sum_{i=1}^{n} \lambda_i - \sum_{i=1}^{k} \lambda_i \right) \left(1 - \sum_{i=1}^{k} p_{ij}^2 \right) \tag{31}$$

which requires only the PCs and eigenvalues retained in the model. The first term on the right hand side in eqn. (31) can be replaced with the total sum of squares, which is equal to the sum over all of the eigenvalues. In chemical processes where many types of sensors with different noise properties are in use, it is probably best to use eqn. (30) to estimate the variance of the residuals for each variable. For applications where the sensors are all closely related, such as in spectroscopy, eqn. (31) may be more appropriate.

Now that the statistical properties of the residuals have been calculated, this information can be used with hypothesis testing to check for failed sensors and changes in the process. This is the general method outlined in [17], and is the basis for many of the methods surveyed in [18–20], which, however, require a complete dynamic model of the process.

It is possible to compare the observed and expected [5] variance of the residuals of individual variables in order to identify changes in the system and its sensors. The standard F-test with the appropriate degrees of freedom may be used. The test will check to see whether

$$s^2_{j_{\text{new}}} / s^2_{j_{\text{old}}} > F_{\nu_{\text{new}},\nu_{\text{old}},\alpha} \tag{32}$$

where

$$\nu_{\text{new}} = m_{\text{new}} - q - 1 \tag{33}$$

$$\nu_{\text{old}} = m_{\text{old}} - q - 1 \tag{34}$$

Here m_{new} and m_{old} are the number of samples in the test data set and training data set, respectively and q is the number of PCs retained in the model. When the inequality in eqn. (32) holds, a change in the variance of the residual has occurred to a confidence level of $1 - \alpha$. The F-test parameters can now be used to set upper and lower limits on the variance of the residuals.

The mean residual should be zero for all the variables. The t-test can be used to detect a shift in the mean away from zero. In this case the hypothesis that the means are equal is to be

[5] i.e., as calculated by eqn. (30) or (31).

tested. Thus the t-test reduces to

$$t_{\nu\ \mathrm{tot}} = \frac{(\overline{x_{\mathrm{old}}} - \overline{x_{\mathrm{new}}})(\nu_{\mathrm{old}} + \nu_{\mathrm{new}})^{0.5}}{(1/\nu_{\mathrm{old}} + 1/\nu_{\mathrm{new}})^{0.5}(\nu_{\mathrm{old}} s^2_{\mathrm{old}} + \nu_{\mathrm{new}} s^2_{\mathrm{new}})^{0.5}} \tag{35}$$

where the degrees of freedom are both one greater than for the case given above. For the purpose of setting limits, the variances can be assumed to be equal to the variance of the residuals of the calibration set as calculated by eqn. (30) or (31). Once the desired confidence level is chosen, it is possible to solve for the difference between the old and new means that is marginally significant.

Note that the tests in (32) and (35) require multiple samples, i.e., a "window" of samples—they do not apply for single observations. In fact, in order to obtain statistical limits, the window has to contain more samples than the number of PCs retained in the model. While it would be possible to observe the residuals from each sample, better sensitivity to changes in the system is obtained by looking at a series of residuals from recent samples. The number of past samples in the series (the width of the window) would be based on the response time and sensitivity desired for the detection scheme. "Wide" windows would allow detection of smaller changes, but would not respond as quickly to changes as "narrow" windows.

In order to determine the "detection power" of the PCA model, the control limits must be converted back to the original units of the data. This is done by first determining the vector of ratios, $\boldsymbol{h}$, of the change in a variable to the change in its residual (with all other variables remaining constant). This is equal to the inverse of the diagonal elements of $(\mathbf{I} - \mathbf{P}_q \mathbf{P}_q^{\mathrm{T}})$, thus

$$\boldsymbol{h} = \left[\mathrm{diag}\left(\mathbf{I} - \mathbf{P}_q \mathbf{P}_q^{\mathrm{T}}\right)\right]^{-1} \tag{36}$$

The PCA detection limits for changes in the residual mean can then be scaled by the $\boldsymbol{h}$ vector to obtain the detection limits in terms of the original variables. In order to obtain the detection limits for changes in the residual variance, the PCA limits must first be converted to standard deviations, then scaled and converted back to variance limits. The result of these scaling operations is that a limit will be established in terms of the original measurement units for detection of biases in senors (arising from sensor drift or a change in the process) and for detection of added noise in sensors (arising from added measurement noise).

EXAMPLE OF THE USE OF PCA IN DYNAMIC SYSTEMS

As an example, we consider a process with $r = 5$ inputs, $n = 5$ states and $p = 10$ measurements The $\mathbf{\Phi}$, Γ, and $\mathbf{C}$ matrices for the process are given below in Tables 1–3. These matrices were generated randomly, although care was taken to assure that the resulting system was asymptotically stable and that the non-zero singular values of the matrices were all relatively large. The $\mathbf{D}$ matrix was zero. Note that $q = n$ in this case.

The example process was driven by white noise of unit variance to produce a $\mathbf{Y}$ matrix consisting of 1000 samples of the 10 outputs. Uncorrelated measurement noise was added to each output sample. The variance of the noise was equal to the variance of the uncorrupted output for each output variable (i.e. the resulting output was 50% deterministic variation and 50% measurement noise). The process outputs

TABLE 1

$\mathbf{\Phi}$ Matrix for example system

$$\begin{bmatrix} 0.3696 & -0.2761 & -0.0582 & -0.6364 & 0.1188 \\ 0.0216 & -0.4511 & -0.2586 & 0.3415 & 0.4932 \\ 0.6204 & -0.0227 & 0.4012 & 0.2988 & 0.0633 \\ -0.2987 & -0.1517 & 0.5948 & -0.1786 & 0.3078 \\ 0.0295 & 0.4772 & -0.0921 & -0.1311 & 0.5786 \end{bmatrix}$$

TABLE 2

Γ Matrix for example system

$$\begin{bmatrix} 0.0955 & -0.6535 & -0.0114 & 0.3726 & 0.0330 \\ -0.3014 & -0.2935 & 0.5805 & -0.2808 & 0.2099 \\ -0.2041 & -0.0979 & -0.4576 & -0.2390 & 0.4931 \\ 0.3669 & 0.1920 & 0.1957 & 0.2874 & 0.6066 \\ 0.5820 & -0.2586 & -0.0619 & -0.4934 & 0.0089 \end{bmatrix}$$

TABLE 3

C Matrix for example system

$$\begin{bmatrix} 0.4219 & -0.1386 & -0.0126 & -0.0922 & -0.0189 \\ 0.0998 & -0.0273 & -0.1266 & -0.4567 & 0.1252 \\ 0.0052 & 0.1546 & 0.05789 & 0.0325 & 0.1544 \\ 0.3851 & 0.0766 & -0.2299 & 0.0466 & 0.3672 \\ 0.0888 & 0.0554 & 0.2707 & 0.2040 & 0.4264 \\ -0.1016 & -0.3428 & 0.3047 & -0.3706 & -0.1408 \\ -0.0620 & -0.6174 & 0.0411 & 0.2417 & 0.1964 \\ 0.0498 & -0.1887 & 0.0267 & -0.2905 & 0.1761 \\ -0.0243 & -0.0041 & 0.0180 & -0.2449 & 0.3680 \\ 0.4857 & -0.0906 & 0.1995 & -0.0216 & -0.2322 \end{bmatrix}$$

TABLE 4

Variance captured by PCA model of example process output

Principal component No.	Eigenvalue	% Variance	%Total variance
1	1.8781	18.7807	18.7807
2	1.6026	16.0255	34.8062
3	1.5282	15.2822	50.0884
4	1.2669	12.6687	62.7571
5	1.2177	12.1765	74.9336
6	0.5906	5.9057	80.8394
7	0.4972	4.9718	85.8111
8	0.4958	4.9578	90.7689
9	0.4715	4.7148	95.4837
10	0.4516	4.5163	100.0000

TABLE 5

PCA Loadings vectors for example system

$$\begin{bmatrix} 0.3650 \\ 0.4742 \\ 0.0061 \\ 0.2664 \\ 0.1229 \\ 0.2120 \\ 0.0566 \\ 0.5346 \\ 0.4509 \\ 0.1360 \end{bmatrix} \begin{bmatrix} -0.0899 \\ 0.2864 \\ -0.4360 \\ -0.4153 \\ -0.5980 \\ 0.3638 \\ -0.0203 \\ 0.1656 \\ -0.1613 \\ -0.0592 \end{bmatrix} \begin{bmatrix} 0.5472 \\ -0.1022 \\ -0.1441 \\ 0.2015 \\ -0.1591 \\ 0.0110 \\ 0.0037 \\ -0.1679 \\ -0.4084 \\ 0.6390 \end{bmatrix} \begin{bmatrix} -0.0084 \\ -0.2648 \\ 0.4595 \\ -0.4155 \\ 0.1821 \\ 0.5580 \\ 0.3583 \\ 0.1277 \\ -0.0538 \\ 0.2323 \end{bmatrix} \begin{bmatrix} 0.0331 \\ -0.1558 \\ -0.3824 \\ 0.1579 \\ 0.0479 \\ -0.1520 \\ 0.8441 \\ 0.1280 \\ -0.0891 \\ -0.2030 \end{bmatrix}$$

were scaled to zero mean and unit variance ("autoscaled") and a PCA model was obtained according to eqn. (24) with $q = 5$. The variance captured by the PCA model is given in Table 4, and the loadings vectors retained in the PCA model are given in Table 5.

Here, the correct number of PCs to retain in the model is known, i.e., 5. In practice this would have to be determined from cross-validation or comparison to the expected ratio of successive eigenvalues for noisy data, as discussed previously.

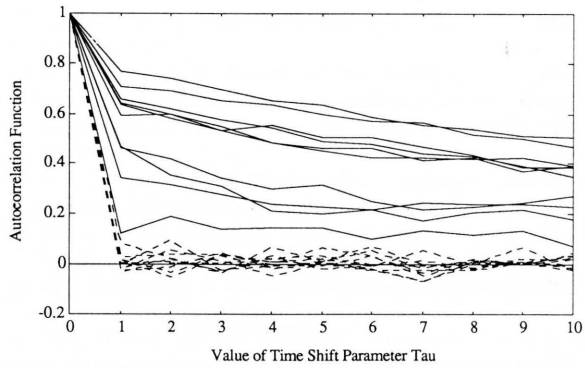

Fig. 1. Autocorrelation function for outputs (solid lines) and residuals (dashed lines) for test data set.

A new data set of 1000 samples was generated by using the same process model and a new input sequence, which in this case was a low-frequency pseudo-random binary sequence (PRBS). The PCA model was applied to the outputs according to eqn. (7), and a residuals matrix was generated. The autocorrelation function was calculated for each output and the residual. These are plotted in Fig. 1, which clearly shows that, while the outputs are correlated in time, the residuals are not, which is the expected result when the $\mathbf{P}_q$ matrix obtained by PCA is an accurate representation of the original $\mathbf{C}$ matrix.

Therefore, the statistical tests in eqns. (30) and (31) can be used to test the residuals for faults and disturbances. As mentioned previously, a sample window width must be chosen, and the desired confidence limits must be set before the test limits can be calculated.

For the example process we will choose, somewhat arbitrarily, a 20 sample window and 99% (i.e. $\alpha = 0.01$) confidence limits. Thus the relevant statistic for detection of changes in the variance of the residuals is $F_{14,994,0.01}$ which equals 2.09, since we have 1000 samples in the original data, 20 in the new test sets, and desire 99% confidence limits. For changes in mean of the residuals $t_{1010} = 2.326$ and it is possible to

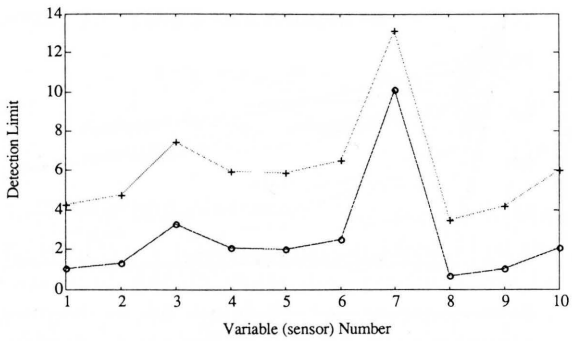

Fig. 2. Detection limits for changes in noise mean (+) and variance (○) in original units.

calculate the change in mean that is marginally significant with eqn. (35), since the variances are known from eqn. (30). The detection limits can now be converted from the residual space back to the original variable space using eqn. (36). The results are shown in Fig. 2, which gives the detection limits for changes in the mean and variance of the measurement noise for each process sensor (variable number) in the original measurement units.

The detection limits shown in Fig. 2 illustrate some important points. Note that variable number 7 has by far the worst detection limits for both mean and variance changes. A review of the loadings given in Table 5, however, shows that it is almost entirely included in the PCA model, i.e., the sum of squared loadings for this variable in the retained PCs is very nearly 1. Variables that act nearly independently tend to be very strongly included in the model and have large detection limits because they are not highly correlated to other variables. Much of the error in such a variable is attributed to variations in the states, and does not appear as a residual. This is easier to envision when one realizes that if a variable was entirely included in the model (i.e., one of the states were defined as being equal to the measured value), its residual would always be zero.

On the other hand, measurements that have little influence on the state estimates also tend to have poor detection limits. The variables with the best detection limits are typically those that are highly correlated with other variables, which tends to make them be included in the model to an intermediate degree. In the situation where a variable does not load into the model at all, its residual variance is equal to its original variance. For a variable like this, standard SPC would work as well as MSPC.

CONCLUSIONS

We have shown when it is appropriate to use PCA to detect faults and upsets in dynamic processes. A particular advantage of the PCA approach (relative, e.g., to the Kalman Filter) is that it provides a convenient way to estimate the **C** matrix in a standard discrete-time, LTI, state–space model. Also, a complete dynamic model of the process is not required.

Proper application of PCA results in residuals that are non-autocorrelated. This greatly simplifies the statistics in fault-detection applications. Standard *F*- and *t*-tests can be used to monitor the residuals for changes in variance (arising from added sensor noise) and mean (arising from sensor bias or process changes), respectively. Other statistical tests, such as the maximum likelihood ratio used by Willsky [18], could also be used (although these are not limited to processes with redundant measurements). We have also shown that detection limits derived for the (scaled) PCA residuals can be used to calculate detection limits in terms of the measurement units of the original variables.

We emphasize that MSPC is most effective when the process has significantly more measurements than states. One could argue that all real systems have an infinite number of states and, therefore, could never have more measurements than states. There are many systems, however, in which a small number of states dominate the dynamic response. The remaining states are associated with transients that decay very quickly relative to the sampling period. For a given process, either an increase in the number of measurements or an increase in the sampling period will usually make MSPC applicable. We contend, therefore, that although MSPC cannot be used in all cases, the number of problems it

can address is significant and is increasing as modern instrumentation systems become more common.

ACKNOWLEDGMENT

This work was partially supported by West Valley Nuclear Services under contract number P.O. 19-23873-N-CG.

REFERENCES

1 B.M. Wise and N.L. Ricker, Feedback Strategies in Multiple Sensor Systems, *AIChE Symp. Ser.*, No. 267, Vol. 85, 1989.
2 B.M. Wise, D.J. Veltkamp, B. Davis, N.L. Ricker and B.R. Kowalski, Principal components analysis for monitoring the West Valley liquid fed ceramic melter, in: R. Post and M. Wades (Eds.), Waste Management '88 Proceedings, Tucson, AZ, 1988.
3 D.J. Veltkamp, B.R. Kowalski, N.L. Ricker and B.M. Wise, Multivariate statistical process control using principal component analysis, *J. Chemometrics*, (1990), submitted.
4 J. Kresta, J.F. MacGregor and T.E. Marlin, Multivariate Statistical Monitoring of process operating performance, *Can. J. Chem. Eng.*, (1990), submitted.
5 N.L. Ricker, Multivariate statistical process control: Analogy to the Kalman filter, *AIChE J.*, (1990), submitted.
6 G. Strang, *Linear Algebra and Its Applications*, Academic Press, New York, NY, 1980.
7 M.A. Sharaf, D.L. Illman and B.R. Kowalski, *Chemometrics*, Wiley, New York, NY, 1986.
8 E.R. Malinowski, Determination of the number of factors and the experimental error in a data matrix, *Anal. Chem.*, 49 (1977) 612.
9 E.R. Malinowski, Theory of the distribution of error eigenvalues resulting from principal component analysis with application to spectroscopic data, *J. Chemometrics*, 1 (1987) 33.
10 K.J. Åstrom and B. Wittenmark, *Computer Controlled Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1984.
11 H. Kwakernaak and R. Sivan, *Linear Optimal Control Systems*, Wiley, New York, NY, 1972.
12 A.P. Sage and C.C. White III, *Optimum Systems Control*, Prentice-Hall, Englewood Cliffs, NJ, 1977.
13 N.L. Ricker, The use of biased least-squares estimators for parameters in discrete-time pulse–response models, *Ind. Eng. Chem. Res.*, 27 (1988) 343–350.
14 J.E.. Jackson and G.S. Mudholkar, Control procedures for residuals associated with principal component analysis, *Technometrics*, 21(3) (1979) 341–349.
15 D.R. Jensen and H. Solomon, A Gaussian approximation to the distribution of a definite quadratic form, *J. Am. Stat. Assoc.*, 67 (340) (1972) 898–902.
16 H. Hotelling, Multivariate quality control illustrated by the air testing of sample bombsights, in: C. Eisenhart, M.W. Hastay and W.A. Wallis (Eds.), *Techniques of Statistical Analysis*, McGraw, New York, NY, 1947, pp. 111–184.
17 R.K. Mehra and J. Peschon, An innovations approach to fault detection and diagnosis in dynamic systems, *Automatica*, 7 (1971) 637–640.
18 A.S. Willsky, A survey of design methods for failure detection in dynamic systems, *Automatica*, 12 (1976) 601–611.
19 R. Isermann, Process fault detection based on modeling and estimation methods, A survey, *Automatica*, 20(4) (1984) 387–409.
20 M. Basseville and A. Benveniste, *Detection of Abrupt Changes in Signals and Dynamical Systems*, Springer-Verlag, Berlin, 1986.